



PhD Thesis Sample

SCENE VISUALIZATION AND RECONSTRUCTION FROM INTERNET PHOTO COLLECTIONS

ABSTRACT

The Internet is currently turning out to be an unprecedented source of different visual information, with literally billions of images that are readily accessible through the different search engines including Flickr and Google Images. These also include thousands of different photographs of almost all famous places, taken from a wide range of viewpoints, and at different times of the day, and even under different weather conditions. This thesis focuses on the problem involved in leveraging these photos in order to come up with new 3D interfaces that makes users possible to virtually explore the world (Garmin, 2011).

One huge challenge is that recreating these 3D sceneries out of photo collections need knowing where every photo was taken. This thesis also introduces the use of new techniques in computer vision, designed to robustly recover these information from different photo collections without the need of GPS and other tools. These techniques are the very first to be presented on Internet images, showing that techniques in 3D reconstruction can be applied successfully to this rich and hugely untapped resource. For this specific problem, scale is a main concern, since Internet collections can be really wide. In this thesis, a new reconstruction algorithm which chooses a small skeletal set of images serve as preprocess. This approach will reduce reconstruction time via an order of magnitude with minimum to zero loss in accuracy or completeness.

INTRODUCTION

A photo is the window to the world. A reliable photo has the capacity to depict an event, or a place in vivid detail and richness, with the capacity to give a viewer a sample of how it feels to be there. However, a single photo comes with a fixed boundary in time and space. Unlike a real window, it is not easy to simply adjust perspectives in order to see things beyond the frame. Included in the photography art is working within established limitations in order to create works of image beauty, capturing it at the right moment and the right composition. However, photographs are captured for all types of reasons on top of being simply artistic. Photos can also communicate and document information about places, people, as well as things in our world, and are also used widely in classified advertisements, commercial advertising and tourism.

At a personal level, images are also used in capturing moments within our lives, so that we can relive memories later, sharing stories with other people. For these types of applications, where the main goal is to capture and share information, the limitations become more significant (Austin, 2013). For example, it may be challenging, even almost impossible to provide a complete understanding of a complex and huge space, or even to document your visit to one of the most popular places with just a single photo.

How are we able to use this rich and vast photo collection to communicate an experience effectively, to provide someone the capacity to move around virtually, exploring a famous landmark, and seeing what it actually looks like at sunrise and sunset, or even at night time, revisiting different events, thus conveying a real understanding of the actual scene. Reality dictates that that vast availability of image collections, as well as the capacity to capture them are not enough in creating the actual experiences. As a matter of fact, with the current tools available, the more images there are available, the harder it becomes to make sense out of them. Most tools for photo browsing treat photos independently, oftentimes visually disconnected.

These tools do not work by exploiting or conveying the richly common structure which exist among the photos. On the contrary, a number of commercial software applications have also started to create some huge photo collections of some urban cityscapes which are organized in a structured way, making it easier to explore underlying scenes. For example, Google Street View has the capacity to stimulate the actual experience of walking down some streets of major cities by showing omnidirectional photos captured at intervals in every city street (Collins, 2012). These web applications and tools, combining with a high visual fidelity of images using a simplified 3D navigation controls, are pretty useful in recreating the experiences of walking around a common street. However, capturing these experiences usually needs a special camera hardware, careful attention to detail and photo capturing process, as well as a time-consuming quality control and post-processing task. Despite the challenges that needs to be faced in order to achieve the goal of this thesis, this project will still be able to succeed as a repository of data, serving as a reference for research, education and for virtual tourism.

References

- Austin, H. (2013). Photography of long scenes using multi-viewpoint panoramas. *SIGGRAPH United*, 12(3), 456-467.
- Collins, H. (2012). Plenoptic stitching: scalable procedure of reconstructing 3D interactive walkthroughs. *Photography Journal*, 23(1), 23-34.
- Garmin, Y. (2011). Spatial navigation for media streaming. *Photography Journal*, 24(3), 34-56.